Best Intentions*

Towards Intention Recognition and Strategy Switching for Mixed-Initiative Recommender Systems

Lorraine McGinty[1] and Barry Smyth[1,2]

[1] Smart Media Institute, University College Dublin, Dublin 4, Ireland
[2] ChangingWorlds, University College Dublin, Dublin 18, Ireland
{firstname.lastname}@ucd.ie

Abstract. Recommender systems are designed to assist users in the search for product and service information and have been successfully deployed in a range of domains, from restaurants to route planning, movies to news. In particular, conversational recommender systems engage the user in an extended recommendation dialog, making suggestions and eliciting user feedback in order to guide the next round of recommendations. This complementary user-system interaction has led to a recent interest in mixed-initiative systems research. In this paper we examine some of the shortcomings of existing case-based conversational recommender systems. In particular, we highlight how a more flexible recommendation strategy, one that responds to intermediate recommendation success and failures, can lead to significant improvements in both the efficiency and quality of recommendation dialogs. We argue that such techniques have a role to play in mixed-initiative recommender systems in the future.

1 Introduction

Conversational recommender systems are good examples of interactive artificial intelligence (IAI) systems [Aha and Munoz-Avila, 2001] where the user engages in a sequence of interactions with the recommender system, providing feedback and additional information about their needs and preferences. The form of feedback is important and distinguishes two important classes of recommendation strategy: those that employ *navigation by asking* and those that employ *navigation by proposing* [Shimazu, 2001]. In the former the user is asked specific questions about specific features within the product space. For instance, in a PC recommender: "How much are you willing to pay?" or "What is your preferred memory size?". In contrast, navigation by proposing avoids asking direct questions and instead tends to present the user with a set of *best guesses* based on their current query, inviting the user to provide feedback by either expressing a simple preference (e.g., "I prefer PC number 2") [McGinty and Smyth, 2002] or a specific feature critique (e.g., "I like PC number 2 but I am looking for more memory.")[Burke et al., 1997].

* The support of the Informatics Research Initiative of Enterprise Ireland is gratefully acknowledged.

In the past, conversational recommender systems have adopted a very rigid structure whereby the user and recommender have fixed roles in a turn-taking conversation [Aha and Munoz-Avila, 2001,Allen, 1999]. However, recently, research from the mixed-initiative systems field has led to the development of more flexible conversational recommendation strategies that can better exploit the relative strengths of human and machine reasoning styles (see for example [Bridge, 2002,Brown and Cox, 1999,McSherry, 2002a]). For instance, mixed-initiative recommenders allow for a more flexible division of responsibility between the user and system [Bridge, 2002]. In addition, one of the important features of a mixed-initiative system is the maintenance of a *shared awareness* with respect to the current state of the human and machine agent(s). In the context of a recommender system this involves helping the user and system to better understand the current recommendation state, both in terms of the user's current and evolving requirements and the system's interpretation of how these requirements map on to the product-space. For example, [McSherry, 2002b] proposes the use of explanations as a way to help the user to better understand the intermediate suggestions of the recommender system.

In this paper we are interested in a related issue, that of *intention recognition* [Allen, 1999,Horvitz and Paek, 1999,Louwerse et al., 2002]. The ability of a recommender system to accurately recognise and respond to the intentions of a user obviously plays an important role in the success of any recommender system. However, we believe that existing recommendation strategies do not fully address this issue. In particular we argue that by blindly following the traditional similarity-assumption – that recommended cases should be maximally similar to the user's current query – recommender systems often misinterpret the user's current intentions, following false leads that result in protracted recommendation dialogs and frustrated or unsatisfied users. We describe a flexible recommendation technique that uses feedback from the user not only to adapt their current query but to also assess recent recommendations were on-target. We propose two different recommendation strategies and a switching mechanism that allows the recommender to switch between them, depending on whether recent recommendations have been on target. This approach has been shown to deliver significant improvements in recommendation efficiency, resulting in shorter recommendation dialogs under a variety of experimental conditions [McGinty and Smyth, 2003b,Smyth and McGinty, 2003]. In this paper we focus on the quality of the recommendation dialogs, arguing that a good dialog should bring the user ever-closer to their ideal recommendation target. We argue that traditional conversational approaches suffer in this regard and present new experiments to show how our approach achieves superior quality.

It should be pointed out at this stage that the system described in this paper is *not* a mixed-initiative recommender system. Nevertheless, the intention recognition approach that we describe, implement and evaluate address relevant challenges in this area and the resulting techniques are potentially useful for true mixed-initiative recommenders.

2 Hill-Climbing and False Leads

One assumption that is often made in conversational recommender systems is the "hill-climbing assumption", where it is expected that each recommender cycle brings the user closer to the appropriate target case/item [Shimazu, 2001]. However, this assumption does not always hold; especially in recommenders operating over complex datasets and relying on critiquing or preference-based feedback. For example, Figure 1 shows the results obtained from graphing the *similarity profiles* of two conversational recommender systems, one that employs critiquing and one that employs preference-based feedback (PBF)(in the Whiskey domain - see Section 4). Each similarity profile is a graph of the similarity of the closest case in a given cycle to the known target case against the number of recommendation cycles. In each graph (ignoring the "+AS" for now) we see a similarity profile that breaks the hill-climbing assumption since many cycles lead to recommendations that are farther from the target case than earlier recommendations.

In general three types of cycle transitions can be distinguished, based on the change in similarity between the best case in a given cycle and the target case; where "best case" refers to the case that is most similar to the target. An *ascent* occurs when the best case in a new cycle is more similar to the target than the best case from the previous cycle. A *plateau* occurs when the best case in the new cycle has the same similarity to the target as that from the previous cycle. Finally, a *descent* occurs when the best case from the new cycle is less similar to the target than the best case from the previous cycle.

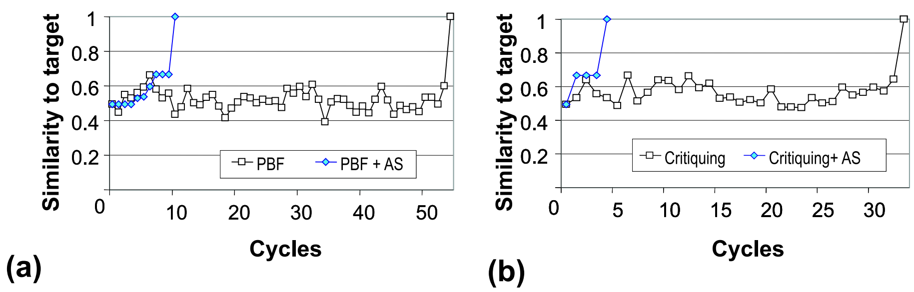

Fig. 1. Similarity profiles for preference-based (a) and critiquing (b) forms of feedback.

In Figure 1(a) we see the similarity profile of a particular query from the Whiskey domain presented to a preference-based recommender system (PBF). Instead of an increasing similarity profile, we find that almost 50% of the cycles are descents for the PBF recommender. For instance, between cycles 7 and 11, target similarity falls from 0.66 to as low as 0.43 as the user is forced to accept poor recommendations. The PBF strategy fluctuates erratically between cycles 8 through to 54. In fact, for the duration of these 46 cycles the recommender conducts an exhaustive search for the target case in what is clearly the wrong region of the recommendation space, given that the user is not presented with any case during this period with a higher similarity to the target than the preference they indicated much earlier on in cycle 7! Obviously, it is unreasonable to expect that any user would interact with a recommender system for this many cycles. This is especially true when there is no evidence of *positive progress* (i.e., the system fails to consistently retrieve cases in each cycle that bring the user closer to the target). In reality, the user will become frustrated and abandon their search unless they see positive progress early on in the recommendation dialogue.

Similar profile trends are characteristic for recommenders using other forms of feedback. For example, Figure 1(b) shows the similarity profile for the same query from the Whiskey domain presented to a recommender system that uses critiquing as it's feedback strategy. Here we find that even this more efficient recommender shows similar cycle transitions in its similarity profile; with ascents (e.g., cycles 1 to 3), descents (e.g., cycles 3 to 6), and plateaus (e.g., cycles 22 to 24). Once again, descents are commonplace. These observations about the similarity characteristic of conversational recommender systems that use preference-based feedback and critiquing as their primary forms of feedback clearly highlight an important problem. Protracted recommendation dialogues must be shortened if such techniques are to be effective in a real-world setting. In the remainder of this paper we propose a technique that is capable of recognising when the recommender is following a false lead, and of guiding the recommender into a more promising region of the product-space.

3 Intention Recognition in Comparison-Based Recommendation

Comparison-based recommendation [McGinty and Smyth, 2002] is a generic framework for conversational recommender systems that emphasises the roles of case selection, user feedback, and query modification during navigation by proposing. It is an iterative recommendation algorithm (Figure 2) that presents the user with a selection of k items as recommendations during each of a sequence of n recommendation cycles. Although initially comparison-based recommendation was proposed as a framework for investigating similarity-based recommenders utilising pure preference-based feedback, it is in fact sufficiently generic to accommodate a range of different recommendation strategies and feedback types [McGinty and Smyth, 2003a,Smyth and McGinty, 2003].

In this work we are particularly interested in the interaction between user and system, whereby both parties contribute to each recommendation cycle. The recommender system contributes a set of recommendations and the user may provide some form of feedback as a response to these recommendations. Ultimately the job of the recommender system is to present the user with a recommendation that is satisfactory and this will only be achieved if a *shared awareness* is developed between system and user with respect to the goals and intentions of each. We are interested in how the feedback provided by the user can be interpreted by the recommender system, not just as a way to update the

```
define Comparison-Based-Recommend(q, CB, k)
  i_p-1, i_p ← null
  do
    R ← ItemRecommend(q, CB, k, i_p, i_p-1)
    i_p ← UserReview(R, CB)
    q ← QueryRevise(q, i_p, R)
    i_p-1 = i_p
  until UserAccepts(i_p)

 define ItemRecommend(q, CB, k, i_p, i_p-1)
    if(i_p != null) && (i_p == i_p-1)
      R ← ReFocus(q, CB, k)
   else
     R ← ReFine(q, CB, k)
   return R

 define QueryRevise(q, i_p, R)
    R′ ← R - {i_p}
    q ← i_p
    return q

define UserReview(R, CB)
  i_p ← user selects preference item from R
  CB ← CB - R
  return i_p

define Refine(q, CB, k)
  CB′ ← sort CB in decreasing order of their sim to q
  R ← top k items in CB′
  return R

define Refocus(q, CB, k, i_p, i_p-1)
  return BoundedGreedySelection(q, CB, k, b)

define BoundedGreedySelection (q, CB, k, b)
  CB' := bk items in CB that are most similar to q
  R := {}
  For j := 1 to k
    sort CB' by Quality(q,i,R) for each case i in CB'
    R  := R + First(CB')
    CB′ := CB′ - First(CB')
  EndFor
  return R
```

Fig. 2. The comparison-based recommendation algorithm with adaptive selection.

user's query [Burke et al., 1997,McGinty and Smyth, 2002], but also as a way to evaluate the relative success of the most recent recommendation cycle.

The essential point is that determining whether the recommender is currently focusing on the correct portion of the recommendation space is important when it comes to deciding on the right recommendation strategy to use in the next recommendation cycle. While traditional similarity-based approaches to recommendation have been used extensively in the past – retrieving the k most similar cases to the query – we believe that such approaches are applicable only when the recommender system is focused on the correct portion of the recommendation space. If the recommender is not well focused then the current query is unlikely to be an accurate or complete representation of the user's true needs. A pure similarity-based approach will have a tendency to over-fit to the assumptions that are made by the recommender about these needs, and it will retrieve cases that match these potentially flawed assumptions. These cases will also have a tendency to be very similar to each other, as well as being similar to the elaborated query, and as such will cover a very limited region of the underlying recommendation space. If incorrect assumptions have been made by the recommender then none of the recommended cases may be suitable, and indeed they may represent poorer recommendations than many of those that have been made previously in the current recommendation session. This is one of the basic reasons that the hill-climbing assumption highlighted above breaks down in practice: similarity-based recommenders over-fit to the local feedback provided by the user and are drawn to nearby cases in the recommendation space when more distant cases may more closely match the user's real intentions.

By presenting a more diverse set of cases the recommender can cover a number of different points in the recommendation space in the hope that one will be a fair match for the user's needs. Thus, in situations where there is evidence that the recommender system is not properly focused, the importance of similarity in recommendation becomes less critical and recommendation diversity becomes more critical. Indeed similar techniques are used by sales assistants in real-world sales dialogs. When a customers needs are unclear the sales assistant will present a diverse set of options to try and focus in on a particular type of product. However, as the user's needs become more refined the sales assistant will tend to switch their recommendation strategy, suggesting products that are as similar as possible to the user's known requirements as they home in on the right region of the product space. These observations motivate the need for a more sophisticated recommendation strategy, one that adapts its response (i.e., it's use of similarity and diversity) depending on whether the it has focused in on the right region of the recommendation space. This is a significant departure for conventional conversational recommender systems and one that it vitally important in the context of mixed-initiative recommender systems. The bottom-line is that an efficient collaboration between user and recommender system can only be achieved if the system is able to recognise when its suggestions are suitable or unsuitable, adjusting its recommendation strategy accordingly.

Adaptive Selection (AS) is such a technique and is described in detail in [McGinty and Smyth, 2003a,Smyth and McGinty, 2003]. It takes advantage of the idea that it is possible to determine if the recommender is correctly focused by recognising if recent recommendations represent an improvement on those made in the previous cycle. This is achieved by making two further modifications to the basic comparison-based recommendation technique. First, instead of making k new recommendations in each new cycle, the current preference case (or the critiqued case) is added to $k-1$ new recommendations. On its own this modification introduces redundancy, in the sense that a previously seen case is repeated in one or more future cycles. However, including the previous preference makes it possible to avoid the problems that ordinarily occur when none of the newly recommended cases are relevant to the user; they can simply reselect the carried preference case instead of being forced to follow a less relevant recommendation.

The essential point is that if the user prefers (or critiques) a case other than the carried preference, then it must be because it is closer to the target, and thus positive progress has been made. In this situation diversity is not warranted and the emphasis should be on similarity in the next recommendation cycle. This corresponds to the **ReFine** method in Figure 2. However, if the user prefers the carried preference case then it suggests that the other $k-1$ cases are less relevant than the carried case, and thus that the recommender has failed to make positive progress towards the target. In this situation two things happen (see **ReFocus** method in Figure 2). First, diversity is introduced into the next recommendation cycle. And secondly, during the selection of the new cases for the next recommendation cycle, the dissimilarity of these candidate cases to the rejected cases is taken into account. The basic idea is to prioritise cases that are not only similar to the query, but also dissimilar to the rejected cases.

The algorithm components in Figure 2 are the modifications needed to implement adaptive selection for use in comparison-based recommendation with preference-based feedback. Once again, directly analogous modifications can be made when critiquing is the form of feedback used [McGinty and Smyth, 2003a].

4 Evaluation

In previous work we evaluated the performance of AS across a range of datasets and for a variety of feedback strategies and found reductions of up to 80% in the number of recommendation cycles needed in preference-based recommender systems [McGinty and Smyth, 2003a,Smyth and McGinty, 2003]. In this section we will re-evaluate the AS technique, this time looking at how it influences the *quality* of recommendation cycles that are produced. In particular we are interested in if it produces recommendation cycles that are more consistent with the hill-climbing assumption, when compared to traditional approaches. For example, the similarity profile graphs (Figure 1) in Section 2 also contain the similarity profiles generated by the AS technique, and it should be clear that these profiles (PBF+AS and Critiquing+AS) are characteristically different from those produced by the standard comparison-based recommendation techniques. For a start the AS profiles are a lot shorter, requiring fewer interactions from the end-user, than those produced by the standard techniques. In Figure 1(a) the AS technique locates the correct target case in cycle 11, compared to cycle 55 for the standard preference-based feedback technique. We also see that the adaptive selection cycles are consistent with the hill-climbing assumption in that none of the cycles are descents. All of the cycles are either ascents (where the user selects a recommended case that is closer to the target case) or plateaus (where the user reselects the preference case that has been carried). For example, in Figure 1(a) cycles 1 to 4 correspond to a plateau; the recommender is not making improved suggestions to the user and the carried preference case is reselected in each cycle. Cycles 5 to 8 correspond to a sequence of ascents where the recommender system has refocused on the correct region of the recommendation space and makes good progress towards the target case.

4.1 Setup

Algorithms We wish to test four different recommender systems divided into two groups of two. One group uses preference-based feedback and one uses critiquing. In each group one of the recommenders employs a standard similarity-approach, in which the k most similar cases to the current query are returned in each cycle, and one uses the AS approach, which employs different recommendation strategies depending on whether the recommender is currently focused in an appropriate part of the recommendation space. Each recommender is implemented using the comparison-based recommendation framework with $k = 3$.

Data-Set and Methodology The *Whiskey* case-base contains a set of 552 cases ([McGinty and Smyth, 2003b]), each describing a particular Scotch whiskey in terms of features such as *distillery*, *age*, *proof*, *sweetness*, *flavour*, *finish* etc. Using a leave-one-out methodology, each case (*base*) in a case-base is temporarily removed and used in two ways. First it serves as the basis for a set of queries constructed by taking random subsets of item features. Second, we select the case that is most similar to the original base. These cases serve as the recommendation *targets* for the experiments. Thus, the base represents the ideal query for a user, the generated query is the initial query that the user provides to the recommender, and the target is the best available case for the user based on their ideal. Each generated query is a test problem for the recommender, and in each recommendation cycle the user's preference is assumed to be the case that is most similar to the known target case. Preference-based feedback or critiquing is applied to this *preference case* as appropriate; in the case of the latter, a random critique is applied to the preferred case in each cycle.

4.2 Results

The results are presented in Figures 3(a-d). For each test query and recommendation technique we analyse the characteristics of the resulting recommendation session, measuring the number of cycle transitions that are ascents, descents, and plateaus. Figures 3(a&c) show the average number of ascents, descents, and plateaus for each recommendation session in the Whiskey domain, comparing pure preference-based and critiquing forms of feedback to their AS counterparts. For example, in Figure 3(a), which corresponds to preference-based feedback, we see that the average recommendation session is made up of 10.3 ascents, 7.1 descents, and 0.08 plateaus; an average session length of 17.48. With PBF+AS the average session length is reduced to 9.4, made up of 4.6 ascents and 4.8 plateaus, but with descents being eliminated entirely. Similar results are found for the Whiskey domain with critiquing as the feedback mechanism (Figure 3(c)).

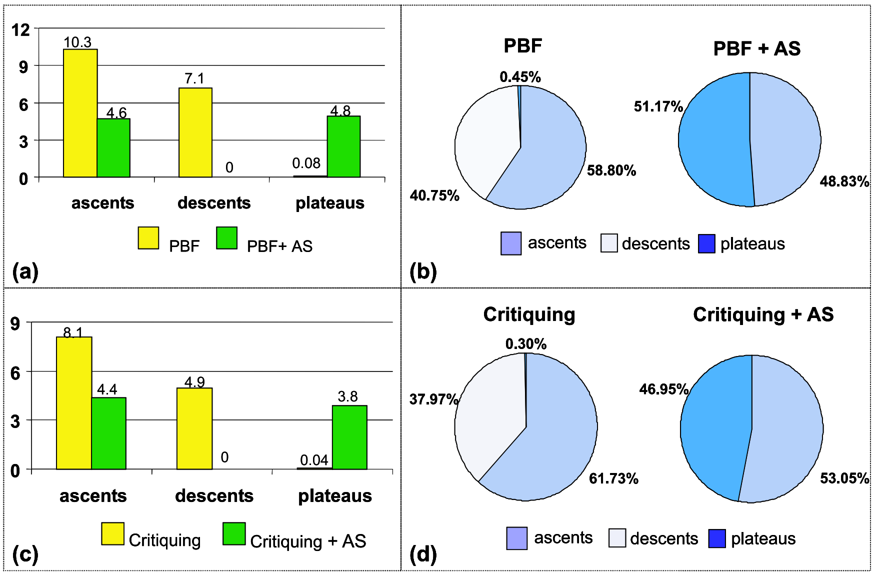

Fig. 3. Evaluation results over the Whiskey dataset.

Interestingly the percentage of cycles that are ascents is lower when AS is used, both when preference-based feedback and critiquing is used. For example, in PBF+AS 49% of cycles are ascents but 58% are ascents in pure PBF. One interpretation of this result concerns changes in seat of control within the system from cycle to cycle. For example, in a traditional conversational recommender system the user is very much in control selecting a *new* case during each cycle as the basis for the next. However, with AS the carrying of the previous preference case into the current cycle gives the user the opportunity to relinquish control in a given cycle by reselecting this carried case, and thus effectively indicating that no improved recommendations have been presented by the recommender. This form of feedback is conceptually very different from the user selecting a new preference case, and indeed signals a change in recommendation strategy as indicated in the previous section. The adaptive section results in Figure 3(b&d) indicate a relatively even sharing of control between the user and system with plateaus occurring between 48% and 53% of the time.

5 Conclusions

Conversational recommender systems generally adopt a fixed recommendation strategy when it comes to making suggestions to users. In this paper we have highlighted how this can lead to protracted recommendation dialogs as the recommender follows false leads, especially when combined with certain forms of feedback, such as preference-based of critiquing. Recently ideas from mixed-initiative systems have informed new approaches to recommender systems that favour a more flexible approach to managing user interactions and guiding recommendations. In this paper we have argued that understanding the intention of the user is critically important and that recommender systems must be capable of recognising when their suggestions are good ones. In turn we have proposed a technique called adaptive selection that is capable of such understanding and of adjusting its recommendation strategy depending on whether recent suggestions are on-target or not. The result is a significant reduction in dialog length and a significant improvement in dialog quality.

Finally, we believe that this work highlights the presence of many other issues to be solved by future recommender systems. For instance, one of the most important features of real-life recommenders, such as sales assistants, is their ability to identify when the customer is ready to buy. In fact, sales may be lost if the assistant does not recognise the customer's readiness to buy; continuing to make suggestions, even good ones, will tend to reduce the likelihood of a successful closure. The point of this example is to highlight yet another "intention" that is worth recognising. In other words, it is important to develop techniques that have the ability to recognise when a user is ready to buy, adjusting the recommender's strategy accordingly; for example, instead of making more product suggestions the recommender may draw the user's attention to a low-cost shipping deal or simply suggest that they move to the checkout process. Thus, in the future then recommender systems will draw on a comprehensive set of strategies and intention recognition techniques that are capable of recognising and coping with different stages in the recommendation process.

References

[Aha and Munoz-Avila, 2001] Aha, D.W. and Munoz-Avila, H. (2001). Interactive case-based reasoning. *Applied Intelligence*, 14(1):7–8.

[Allen, 1999] Allen, J. (1999). Mixed-Initiative Interaction. *IEEE Intelligent Systems, Trends and Controversies*, 14(5):14–16.

[Bridge, 2002] Bridge, D. (2002). Towards Conversational Recommender Systems: A Dialogue Grammar Approach. In Aha, D.W., editor, *Proceedings of the Mixed-Initiative Workshop on CBR at the European Conference on Case-Based Reasoning (ECCBR-02).*, pages 8–22. The Robert Gordon University. Aberdeen, Scotland.

[Brown and Cox, 1999] Brown, S. and Cox, M. (1999). Planning for Information Visualization in Mixed-Initiative Systems. In *Proceedings of the AAAI-99 Workshop on Mixed-Initiative Intelligence.* AAAI. Orlando, Florida, USA.

[Burke et al., 1997] Burke, R., Hammond, K., and Young, B. (1997). The FindMe Approach to Assisted Browsing. *Journal of IEEE Expert*, 12(4):32–40.

[Horvitz and Paek, 1999] Horvitz, E. and Paek, T. (1999). A Computational Architecture for Conversation. In *Proceedings of the Seventh International Conference on User Modeling.*, pages 201–210. Springer Wien. Banff, Canada.

[Louwerse et al., 2002] Louwerse, M., Graesser, A., and Olney, A. (2002). Good computational manners: Mixed-initiative dialog in conversational agents. In Miller, C., editor, *Etiquette for Human-Computer Work. Papers from the 2002 Fall Symposium.*, pages 71–76.

[McGinty and Smyth, 2002] McGinty, L. and Smyth, B. (2002). Comparison-Based Recommendation. In Craw, S., editor, *Proceedings of the Sixth European Conference on Case-Based Reasoning (ECCBR-02)*, pages 575–589. Springer. Scotland.

[McGinty and Smyth, 2003a] McGinty, L. and Smyth, B. (2003a). The Role of Diversity in Conversational Systems. In Bridge, D. and Ashley, K., editors, *Proceedings of the Fifth International Conference on Case-Based Reasoning (ICCBR-03).* Springer. Troindheim, Norway.

[McGinty and Smyth, 2003b] McGinty, L. and Smyth, B. (2003b). Tweaking Critiquing. In *Proceedings of the Workshop on Personalization and Web Techniques at the International Joint Conference on Artificial Intelligence (IJCAI-03).* Morgan-Kaufmann. Acapulco, Mexico.

[McSherry, 2002a] McSherry, D. (2002a). Mixed Initiative Dialogue in Case-Based Reasoning. In Aha, D.W., editor, *Proceedings of the Mixed-Initiative Workshop on CBR at the European Conference on Case-Based Reasoning (ECCBR-02).*, pages 1–8. The Robert Gordon University. Aberdeen, Scotland.

[McSherry, 2002b] McSherry, D. (2002b). Recommendation Engineering. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI-02).*, pages 86–90. Lyon, France.

[Shimazu, 2001] Shimazu, H. (2001). ExpertClerk : Navigating Shoppers' Buying Process with the Combination of Asking and Proposing. In Nebel, B., editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, pages 1443–1448. Morgan Kaufmann. Seattle, Washington, USA.

[Smyth and McGinty, 2003] Smyth, B. and McGinty, L. (2003). The Power of Suggestion. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-03).* Morgan-Kaufmann. Acapulco, Mexico.